

Serial No. 09/728,689

Amdt. Dated ~~25~~ July 2006

Reply to Notice of Non-Compliant Amendment of 31 January 2005

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Cancel claims 1-42 without prejudice.

Add the following new claims 43-46, as follows.

Listing of Claims:

1. (canceled)

2. (canceled)

3. (canceled)

4. (canceled)

5. (canceled)

6. (canceled)

7. (canceled)

8. (canceled)

9. (canceled)

10. (canceled)

11. (canceled)

12. (canceled)

Serial No. **09/728,689**

Amdt. Dated **28** July 2006

Reply to Notice of Non-Compliant Amendment of 31 January 2005

13. (canceled)

14. (canceled)

15. (canceled)

16. (canceled)

17. (canceled)

18. (canceled)

19. (canceled)

20. (canceled)

21. (canceled)

22. (canceled)

23. (canceled)

24. (canceled)

25. (canceled)

26. (canceled)

27. (canceled)

Serial No. **09/728,689**

Amdt. Dated ~~25~~ July 2006

Reply to Notice of Non-Compliant Amendment of 31 January 2005

28. (canceled)

29. (canceled)

30. (canceled)

31. (canceled)

32. (canceled)

33. (canceled)

34. (canceled)

35. (canceled)

36. (canceled)

37. (canceled)

38. (canceled)

39. (canceled)

40. (canceled)

41. (canceled)

42. (canceled)

43. **(new)** A method for extracting records from a structured text in a computer system, comprising:

identifying potential locations of values of record fields in the text by identifying locations in the text of items in lists of known potential values for record fields,

identifying a region of interest in the text by applying multiple candidate region partitioners, evaluating each to measure how well it isolates a region with a high density and a high amount of potential locations of values of record fields, selecting one that measures best, and applying it to produce a region of interest,

segmenting the region of interest into record regions that each contain data for a single record by applying multiple candidate segmenters, evaluating each to measure how well it segments into regions such that each region has one field value per record field and such that different regions have similar numbers of field values for each record field, selecting one that measures best, applying it to produce record regions, extracting field values from record regions by identifying most likely locations of field values for each record field in each record region, and

outputting records composed of extracted field values for record fields.

44. **(new)** The method of claim 43, with the addition of:

identifying potential locations of values of record fields in the text by identifying locations in the text of patterns of potential values for record fields.

45. **(new)** The method of claim 43, with the addition of:

identifying potential locations of values of record fields in the text by
identifying locations in the text of numbers in ranges that are potential
values for record fields.

46. (new) An apparatus for extracting data from a file, comprising a
computer and a computer program, performed by the computer, for:

identifying potential locations of values of record fields in the text by
identifying locations in the text of items in lists of known potential values
for record fields,

identifying a region of interest in the text by applying multiple
candidate region partitioners, evaluating each to measure how well it
isolates a region with a high density and a high amount of potential
locations of values of record fields, selecting one that measures best, and
applying it to produce a region of interest,

segmenting the region of interest into record regions that each
contain data for a single record by applying multiple candidate
segmenters, evaluating each to measure how well it segments into
regions such that each region has one field value per record field and
such that different regions have similar numbers of field values for each
record field, selecting one that measures best, applying it to produce
record regions,

extracting field values from record regions by identifying most likely
locations of field values for each record field in each record region, and

outputting records composed of extracted field values for record
fields.